eCV: Enhanced coefficient of variation and IDR extensions for reproducibility assessment of high-throughput experiments with multiple replicates.

Agustin Gonzalez-Reymundez[1], Kylie Shen[1], Wayne Doyle[1], Sichong Peng[1], Kasey Hutt[1], Stephanie Bruns[1]

[1]Eclipsebio, San Diego, CA.

ABSTRACT

Motivation

Reproducibility assessment is essential in extracting reliable scientific insights from high-throughput experiments. Inconsistency between technical replicates poses a challenge, particularly clear in next generation sequencing technologies based on immunoprecipitations, where the assessment of reproducibility in peak identification is a critical analytical step. While the Irreproducibility Discovery Rate (IDR) method has been instrumental in assessing reproducibility, its standard implementation is constrained to handling only two replicates. In the current era of steadily growing sample sizes, eased by multiplexing and reduced sequencing costs, highly performing methods that handle any number of replicates are desirable.

Results

We introduce three novel methods for reproducibility assessment in high-throughput data that handle an arbitrary number of replicates. The first, general IDR (gIDR), extends the standard IDR by adapting its Expectation-Maximization (EM) algorithm to handle distributions of any dimensions dictated by the number of replicates. The second, meta-

IDR (mIDR), employs a meta-analysis approach, calculating local IDR scores for all pairs of replicates and combining them using standard probability rules. The third method introduces an "enhanced" Coefficient of Variation (eCV), ranking features based on intensity and variability, using a parametric bootstrap approach to obtain an index analogous to local IDR. Comparative analysis with traditional IDR in simulated and experimental data reveals the heightened performance of the proposed multivariate alternatives under varying scenarios, thereby addressing the critical challenge of reproducibility assessment in contemporary high-throughput experiments.

Availability and implementation

The described methods are implemented as an R package: https://github.com/eclipsebio/eCV

Contact

info@eclipsebio.com

1 INTRODUCTION

The reproducibility of high-throughput experiments is essential in assuring high-quality results and valid scientific insights (Nicholson and Holmes, 2017). Lack of consistency within technical replicates, whether coming from heightened variability or diminished mean intensities, can undermine the reliability of the finding. For example, the lack of reproducibility between peaks in separated ChIP-seq libraries has been recognized as one of the most important pitfalls of the technology (Mundade *et al.*, 2014; Landt *et al.*, 2012). Li et al. (2011) introduced the Irreproducibility Discovery Rate (IDR) to measure the reproducibility of high-throughput data. This method robustly models the distribution of omic feature intensity along each technical replicate using a copula model (CM), and models conveniently the joint distribution of omic features in each sample replicate based only on the marginal distributions (Genest and Nešlehová, 2013).

Within the IDR framework, features are transformed and assumed to come from a mixture of two normal distributions, each representing one of two possible states: reproducible or irreproducible. An expectation-maximization (EM) algorithm (Dempster *et al.*, 1977) is then applied to estimate the parameters of this mixture, including the mean and variance of each normal distribution along with the mixing probability. Upon convergence, the likelihood of an omic feature belonging to the reproducible state is used to compute a "local" IDR score. To account for situations where the distribution of omic-feature values is unknown or too complex to approximate via elemental distributions (e.g., normal, negative binomial, gamma), a "semiparametric" CM approach is used, where marginal empirical cumulative distributions of each replicate are fed into the CM (Tsukahara, 2005). The actual IDR is calculated analogously to an adaptation step-up procedure to control

61 the false discovery rate (FDR) (Sun and Cai, 2007; Li *et al.*, 2011). After the model is
62 fitted, features with small IDR scores are taken as "reproducible," while features with
63 higher IDR scores are excluded as "irreproducible."

64 IDR has shown robust performance in extensive simulations (Li et al., 2011) and has also
65 been adopted by the ENCODE project as the gold standard to assess the reproducibility
66 of ChIP-seq data (Landt *et al.*, 2012). Nevertheless, the classical implementation of IDR
67 has the limitation of being unable to handle more than two replicates. The reason for this
68 constraint imposed on IDR is the conclusion from early ChiP-seq experiments that more
69 than two replicates would increase the assay's cost while not significantly improving
70 performance values, such as site discovery (Rozowsky *et al.*, 2009). However,
71 multiplexing and the continuous drop in sequencing costs make it more affordable for
72 researchers to increase sample size (Chen *et al.*, 2012). Larger sample sizes can improve
73 estimates of the variability associated with noise and artifacts effects and increase
74 statistical power (Li *et al.*, 2017).

75 Several alternatives to measure reproducibility in an arbitrary number of replicates have
76 been proposed, including pooling of all replicates (Chen *et al.*, 2011; Landt *et al.*, 2012),
77 selection of the best pairs of replicates (Revilla-i-Domingo *et al.*, 2012), or "majority rule"
78 (Haecker *et al.*, 2012). However, as Yang et al. (2014) discussed, the quality of these
79 data-combination strategies may be susceptible to outliers, diminish signal-to-noise
80 ratios, or compromise subsequent quantitative comparisons among samples.

81 To overcome these limitations, reproducibility assessment methods exploiting the true
82 multivariate nature of multi-replicate experiments are needed. Here, we present three

83 novel approaches to generalize the reproducibility assessment of high throughput data
84 on an arbitrary number of replicates. The first one, general IDR (gIDR), expands the
85 classic IDR by extending its EM approach to handle distributions of arbitrary dimensions
86 dictated by the number of replicates. The second one called meta-IDR (mIDR), resembles
87 a meta-analysis procedure where local IDR scores are calculated for every pair of
88 replicates and combined using standard probability rules. The third method is based on
89 the coefficient of variation rather than IDR. It ranks features in terms of their intensity and
90 variability and uses a parametric bootstrap approach to estimate an index analogous to
91 the local IDR. This method is an "enhanced" Coefficient of Variation (eCV).

92 All three methods presented here were compared against traditional IDR in simulated and
93 experimental data of varying numbers of replicates. Simulated data followed three
94 scenarios used in the original IDR publication, demonstrating situations where the
95 reproducible features are a) many and strongly correlated, b) a few and weakly correlated,
96 and c) even fewer but strongly correlated. The methods were also compared in miR-
97 eCLIP (Manakov *et al.*, 2022) and RBP-eCLIP (Van Nostrand *et al.*, 2016) data. We show
98 that IDR is outperformed by multivariate alternatives, particularly by gIDR when many
99 highly correlated genuine features are present (such as miR-eCLIP, where RNAs form
100 chimeras with highly abundant miRNAs). Methods eCV and mIDR, on the other hand,
101 performed well when a small number of highly correlated reproducible features were
102 present (e.g., reads clustering at transcription factor or RNA binding protein binding sites).
103 In addition, the eCV method resulted in the most stable performance across all scenarios
104 when a liberal threshold on the probability is used (50%). All approaches are implemented
105 in the R package *eCV*.

2 MATERIALS AND METHODS

2.1. Data sets

2.1.1 Simulated data sets

Simulated data was generated using an extension of the copula mixture model presented in Li et al. (2011), assuming 1,000 omic features with values coming from two, three, and four replicates, respectively. In each simulation, the values of the i-th feature were defined as $y_{ij} = -\log\left(\frac{1}{1+e^{x_{ij}}}\right)$. The vector holding the values of feature i, $\boldsymbol{x}_i = [x_{i1}, \dots, x_{ir}]$, is assumed as coming from a mixture of two multivariate normal distributions (ϕ) of dimension r:

$$\boldsymbol{x}_i \sim \phi_0(\mu_0\mathbf{1}, \boldsymbol{\Sigma}_0)^{(1-K_i)}\phi_1(\mu_1\mathbf{1}, \boldsymbol{\Sigma}_1)^{K_i} \quad (1)$$

Such as

$$\phi_k \propto |\boldsymbol{\Sigma}_k^{-1}|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i - \mu_k\mathbf{1})(\boldsymbol{x}_i - \mu_k\mathbf{1})^T)\right) \quad (2)$$

Where the dummy variable K_i, standing for the reproducible or irreproducible classes, was sampled from a Bernoulli variable K_i with probability π.

$$K_i \sim \mathrm{Ber}(\pi) \quad (3)$$

Parameters μ_k and $\boldsymbol{\Sigma}_k$ represent a mean scalar and an $r\mathrm{x}r$ variance-covariance matrix, respectively. The quantity $\mathbf{1}$ was a vector of ones of length r. When $K_i = 0$, irreproducible features values were sampled from ϕ_0, with $\mu_0 = 0$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}$ being the identity matrix (i.e., a matrix with all zeros, except for ones on the diagonal). Whenever $K_i = 1$, features values were sampled from ϕ_1.

In all cases, features were assumed to have unit variance, the same mean intensity (μ_1= 2.5) and the same correlation values. This implied following a compound-symmetric

model for the variance-covariance matrix $\mathbf{\Sigma}_1 = \sigma_1^2(1-\rho_1)\left[\mathbf{I} + \frac{\rho_1}{1-\rho_1}\mathbf{J}\right]$, where σ_1^2 and ρ_1 are the reproducible class variance and correlation, respectively, $\mathbf{I}$ is an $r\mathrm{x}r$ identity matrix and $\mathbf{J}$ is an $r\mathrm{x}r$ matrix with all ones.

A summary of each parameter's value defining each simulation scenario is presented in Table 1. Scenario 1 corresponded to a case where most features came from the reproducible class ($\pi = 0.65$) and were highly correlated ($\rho_1 = 0.84$). Scenario 2 corresponded to a case with few ($\pi = 0.25$) lowly correlated ($\rho_1 = 0.40$) reproducible features. Scenario 3 represented cases with even fewer ($\pi = 0.05$) but highly correlated ($\rho_1 = 0.84$) reproducible features. The instructions to generate the simulated data can be found in Supplementary File S1.

Table 1: Parameters values defining each simulation scenario. Each simulated scenario (**Scenario 1-3**) was defined by assigning values to population parameters (**Parameter values**); proportion (π), mean intensity (μ_1), variance (σ^2_1), and correlation (ρ_1) of reproducible features.

Parameters values	Scenario 1	Scenario 2	Scenario 3
π	0.65	0.25	0.05
μ_1	2.50	2.50	2.50
σ^2_1	1.00	1.00	1.00
ρ_1	0.84	0.40	0.84

140 *2.1.2 Experimental data sets*

141 Experimental data came from two miR-eCLIP (Manakov *et al.*, 2022) and one RBFOX2
142 eCLIP (Van Nostrand *et al.*, 2016) data sets. The first scenario corresponded to miR-
143 eCLIP on HEK293T cells transfected with two microRNAs (miRNA), *hsa-miR-1-3p* (miR-
144 1) and *hsa-miR-124-3p* (miR-124). The data was arranged as intensity of messenger
145 RNA and miRNA interactions (mRNA:miRNA), in the form of chimeric reads along two
146 technical replicates. In this scenario, chimeric peaks annotated with the transfected
147 miRNA (i.e., mRNAs forming chimeras with either one of the exogenous miRNAs) were
148 used to measure performance. In this context, peaks annotated with the transfected
149 miRNA and classified as reproducible, were taken as "true positives". Meanwhile,
150 endogenous mRNA:miRNA chimeras, classified as irreproducible, were taken as "true
151 negatives". We justified this choice by the fact that most endogenous chimeras in this
152 data set had significantly lower intensities than the ones involving transfected miRNAs
153 (Supplementary File S2). Since only two replicates were available ($r = 2$), the only
154 relevant comparisons were between IDR and eCV, since both gIDR and mIDR converge
155 exactly to IDR when $r = 2$.

156 The second experimental scenario also consisted of miR-eCLIP data, but this time on
157 wild-type K562 cells, using two, three, and four replicates. Since all chimeras involved
158 endogenous RNAs, we used a different definition of "truth". In this scenario, true positives
159 were defined by chimeras with positive seed match on the 3' untranslated region (UTR).
160 A positive seed match (i.e., high-quality alignment between the seed region of a miRNA
161 and the complementary sequence of its target mRNAs) on the 3'UTR of a gene (the region
162 of the targeted gene most likely to generate stable and effective RNA duplexes) show a

163 potential regulatory interaction (Gu *et al.*, 2009). Seed matching analysis was performed
164 with *scanMiR* (v4.2.1) and *scanMiRData* (v4.2.1) (Soutschek *et al.*, 2022).
165 The third experimental scenario comprised RBFOX2-eCLIP data, organized as peaks
166 generated around clusters created by crosslinking and immunoprecipitating RBFOX2
167 across two, three, and four technical replicates from wild-type HEK293T cells. Peaks
168 enriched with the RBFOX2 motif were detected using *HOMER* (Heinz *et al.*, 2010). Peaks
169 enriched with the RBFOX2 motif and mapping onto introns or the 3'UTR of targeted
170 transcripts were taken as proxies of "truly reproducible" features. Every other type of peak
171 was considered "truly irreproducible." A summary of the characteristics of each
172 experimental scenario is provided in Table 2.
173 **Table 2: Characteristics of each experimental data scenario**.

Characteristics	**Scenario 1**	**Scenario 2**	**Scenario 3**
Technology	miR-eCLIP	miR-eCLIP	RBP-eCLIP
Target of immune precipitation	AGO2	AGO2	RBFOX2
Experimental condition	Transfections with miR-1 and miR-124	Wild type	Wild type
Cell type	HEK293T	K562	HEK293T
Feature type	Peaks formed by chimeric reads	Peaks formed by chimeric reads	Peaks formed by IP reads normalized by input reads

Number of replicates (r)	Two	Four	Four
Proxy for truly reproducible features (K_i = 1)	Chimeric peaks annotated with the transfected miRNA	Chimeric peaks on 3' UTR with positive seed matching	Peaks on introns or 3' UTR with RBFOX2 motif

174

175 The protocol details for both technologies can be found in Manakov et al. (2022) and Van
176 Nostrand et al. (2016), respectively. Briefly, miR-eCLIP is an extension of the standard
177 eCLIP protocol using an anti-AGO2 antibody and modified to enable chimeric ligation of
178 miRNA and target mRNA. RBFOX2 RBP-eCLIP (RBFOX2-eCLIP), on the other hand,
179 was obtained by applying RBP-eCLIP against RBFOX2, an RNA-binding protein with a
180 widespread role in several cellular mechanisms and tissue-specific effects (Arya *et al.*,
181 2014). RBP-eCLIP extends the single-end eCLIP protocol (Van Nostrand *et al.*, 2016) by
182 optimizing several aspects of the original technique, like cell input requirement, UV
183 crosslinking settings, and the lysis protocol.
184 The resulting sequencing data went through multiple steps of preprocessing and then
185 clusters of aligned reads were found using CLIPper (v2.0.1). For the miR-eCLIP data sets,
186 each cluster was annotated with the names of miRNAs responsible for that target. For the
187 RBFOX2 eCLIP, the IP levels were normalized with input levels to calculate fold
188 enrichments, and p-values were calculated using a Yates' Chi-Square test (or Fisher
189 Exact Test if the observed or expected read number was below 5). In all datasets, peaks
190 were annotated using transcript information from GENCODE release 41 (GRCh38.p13)

191 (Frankish *et al.*, 2022) with the following priority hierarchy to define the final annotation of
192 overlapping features: protein-coding transcript (CDS, UTRs, intron), followed by non-
193 coding transcripts (e.g., exon, intron).

194 **2.2 Reproducibility assessment methods: models and algorithms**

195 *2.2.1 Estimation of mixture copula model parameters*

196 Suppose $\mathbf{X} = (x_{ij})_{i=1,\dots,n}^{j=1,\dots,r}$ is a matrix with positive values standing for significance or
197 intensity of n omic features in r technical replicates. As in standard IDR, gIDR and mIDR,
198 used a copula mixture model on $\mathbf{X}$. In the case of mIDR, the model fits in every pairwise
199 combination of replicates. The derivation of the copula model and parameters estimation
200 is as follows. First, feature values are transformed according to the probability integral
201 transform (Savits, 1994) and re-scaled to avoid infinities:

$$u_{ij} \equiv \frac{n}{n+1}\hat{F}_j(x_{ij}). \tag{4}$$

202 Where $\hat{F}_j$ is the empirical cumulative probability distribution (ECDF) of omic feature values
203 across replicate j.

204 Second, the vector of model parameter values is initialized $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)} = \left(\pi_1^{(0)},\ \mu_1^{(0)},\ \boldsymbol{\Sigma}_1^{(0)}\right)$.
205 Where $\boldsymbol{\Sigma}_1 = \sigma_1^2(1-\rho_1)\left[\mathbf{I} + \frac{\rho_1}{1-\rho_1}\mathbf{J}\right]$, σ_1^2 is the reproducible class variance, ρ_1 is the
206 reproducible class correlation, $\mathbf{I}$ is a rxr matrix with ones in the diagonal and zero
207 otherwise and $\mathbf{J}$ is rxr matrix with all ones. In the case of the irreproducible class, $\boldsymbol{\Sigma}_0 = \mathbf{I}$.

208 Third, $\mathbf{X}$ is turned into "pseudo data" $\mathbf{Z} = (z_{ij})_{i=1,\dots,n}^{j=1,\dots,r}$, where $z_{ij} = G^{-1}(u_{ij}|\boldsymbol{\theta})$, with

$$u_{ij} \equiv G(z_{ij}) = \frac{\pi_1}{\sigma_1}\Phi\left(\frac{z_{ij} - \mu_1}{\sigma_1}\right) + \Phi(z_{ij}) \tag{5}$$

and Φ is the univariate standard normal cumulative density function. The value of G^{-1} is obtained by calculating G on regular grid of 1,000 u_{ij} values followed by linear interpolation as in Li et al. (2011).

Then, an EM algorithm on the augmented pseudo data $\boldsymbol{y}_i = (K_i, \boldsymbol{z}_i)$ is used to obtain $\boldsymbol{\theta}^{(t)}$ (i.e., the estimates of parameter values at iteration t) where $\boldsymbol{z}_i = [z_{i1}, \dots, z_{ir}]$. After setting $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$**,** the last two steps are repeated until convergence or a maximum value to t (t_{MAX}) are reached.

2.2.2 gIDR

The principal difference between IDR and gIDR is that the latter incorporates information from all replicates by using the following complete log-pseudolikelihood formula:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left[\begin{matrix}(1 - K_i)(\log(1-\pi) + \log(\phi_0(\boldsymbol{z}_i|\mu_0, \boldsymbol{\Sigma}_0))) + \\ K_i(\log(\pi) + \log(\phi_1(\boldsymbol{z}_i|\mu_1, \boldsymbol{\Sigma}_1)))\end{matrix}\right]. \tag{6}$$

where ϕ_k is the probability density function of a “General” Multivariate Normal distribution, with dimension equal to the number of replicates r (see equation (2))

Since the model is in essence a mixture of two normal distributions on the pseudo data, standard EM formulas apply (Xu *et al.*, 2016). Hence, for the Expectation step (*E-step*) we have

$$K_i^{(t+1)} = E\left(K_i|z_{i1}, \dots, z_{ir}, \boldsymbol{\theta}^{(t)}\right) = \frac{\pi_1^{(t)}\phi_1\left(z_{i1},\dots,z_{ir}|\mu_1^{(t)}, \boldsymbol{\Sigma}_1{}^{(t)}\right)}{\sum_{k=0}^{1}\pi_k^{(t)}\phi_k\left(z_{i1},\dots,z_{ir}|\mu_k^{(t)}, \boldsymbol{\Sigma}_k{}^{(t)}\right)}, \tag{7}$$

For the Maximization step (*M-step*), we have:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \pi_1} = \sum_{i=1}^{n} \frac{(1-K_i)}{1-\pi_1} - \frac{K_i}{\pi_1} = 0, \tag{8}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \mu_1} = \sum_{i=1}^{n} K_i \, \boldsymbol{z}_i^T \boldsymbol{\Sigma}_1^{-1} - \mu_1 \mathbf{1}^T \boldsymbol{\Sigma}_1^{-1} \sum_{i=1}^{n} K_i = 0, \tag{9}$$

and

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}_1^{-1}} = \boldsymbol{\Sigma}_1 \sum_{i=1}^{n} K_i - \sum_{i=1}^{n} K_i \, (\boldsymbol{z}_i - \mu_k \mathbf{1})(\boldsymbol{z}_i - \mu_k \mathbf{1})^T = 0 \tag{7}$$

Where in equation (10) we used a) the derivatives of the logarithm of a matrix's determinant with respect to $\boldsymbol{\Sigma}_1^{-1}$ and b) the derivative of the trace of the product between a matrix and $\boldsymbol{\Sigma}_1^{-1}$ with respect to $\boldsymbol{\Sigma}_1^{-1}$ (Dwyer, 1967).

The maximum likelihood estimates of each parameter in the $(t+1)$-th iteration are given by the following expressions:

$$\pi_1^{(t+1)} = \sum_{i=1}^{n} \frac{K_i^{(t+1)}}{n}, \tag{11}$$

$$\mu_1^{(t+1)} = \frac{\sum_{i=1}^{n} K_i^{(t+1)} \sum_{j=1}^{r} z_{ij}}{r \sum_{i=1}^{n} K_i^{(t+1)}} \tag{12}$$

and

$$\boldsymbol{\Sigma}_1^{(t+1)} = \frac{\sum_{i=1}^{n} K_i^{(t+1)} \left(\boldsymbol{z}_i - \mu_1^{(t+1)} \mathbf{1}\right)\left(\boldsymbol{z}_i - \mu_1^{(t+1)} \mathbf{1}\right)^T}{\sum_{i=1}^{n} K_i^{(t+1)}}. \tag{13}$$

Finally, since $\boldsymbol{\Sigma}_1 = \sigma_1^2 (1-\rho_1)\left[\mathbf{I} + \frac{\rho_1}{1-\rho_1}\mathbf{J}\right]$, we have:

$$(\sigma_1^2)^{(t+1)} = \frac{\mathrm{tr}\left(\boldsymbol{\Sigma}_1^{(t+1)}\right)}{r} = \frac{\sum_{i=1}^{n} K_i^{(t+1)} \sum_{j=1}^{r} \left(z_{ij} - \mu_1^{(t+1)}\right)^2}{r \sum_{i=1}^{n} K_i^{(t+1)}} \tag{8}$$

and

$$\rho_1^{(t+1)} = \frac{\mathrm{trU}\left(\boldsymbol{\Sigma}_1^{(t+1)}\right)}{(\sigma_1^2)^{(t+1)} \binom{r}{2}} = \frac{2 \sum_{i=1}^{n} K_i^{(t+1)} \sum_{j \neq l} \left(z_{ij} - \mu_1^{(t+1)}\right)\left(z_{il} - \mu_1^{(t+1)}\right)}{(\sigma_1^2)^{(t+1)} r(r-1) \sum_{i=1}^{n} K_i^{(t+1)}}, \tag{9}$$

where tr and trU are the trace and upper-triangular trace of a matrix, respectively, and $\binom{r}{2}$ is the number r replicates combinations in two. Finally, the local IDR is approximated as $\Pr(K_i = 0 | \boldsymbol{\theta}, \mathbf{X}) \approx 1 - K_i^{(t_{MAX})}$, where t_{MAX} is the number of iterations at convergence. Notice that when $r = 2$, the *E-step* and *M-step* formulas above reduce to ones derived for standard IDR (see Li *et al*, 2011 and supplementary materials there).

2.2.3 mIDR

Method mIDR is derived using a simple heuristics method that resembles a meta-analysis score (Shelby and Vaske, 2008). First, we perform standard IDR in each one of every

pairwise combination of replicates. Then, a meta score is calculated to get the probability of a feature being irreproducible in at least one of the $\binom{r}{2}$ comparisons:

$$\Pr(K_i = 0|\boldsymbol{\theta}, \mathbf{X}) \approx 1 - \prod_{j \neq l} \Pr\left(K_i = 1 \middle| \boldsymbol{\theta}, x_{ij}, x_{il}\right) \tag{10}$$

2.2.4 eCV

The eCV method assumes omic feature values across replicates $\boldsymbol{y}_i = [y_{i1}, y_{i2}, \dots, y_{ir}]$ are positive and approximately Normal. If x_{ij} stands for the number of aligned sequencings reads or another measurement of the feature's intensity, with distribution approximately log-Normal, taking $y_{ij} = \log(x_{ij} + \in)$, where $\in$ is an added positive value that prevents infinite values, yields the same distribution as in (1).

The estimated eCV is calculated as:

$$\widehat{eCV}_i = \frac{\left|\hat{\sigma}_i^2 - \hat{\mu}_i^2\right|}{\hat{\mu}_i}, \tag{11}$$

where $\hat{\sigma}_i^2 = \frac{1}{(r-1)} \sum_{j=1}^{r} \left(y_{ij} - \hat{\mu}_i\right)^2$ and $\hat{\mu}_i = \frac{1}{r} \sum_{j=1}^{r} y_{ij}$. Smaller values of $\widehat{eCV}_i$ would show reproducibility and vice versa.

To make inferences on the reproducibility of feature i, we first compute the following expression

$$\Pr\left(eCV_i \leq \widehat{eCV}_i | \boldsymbol{y}_i, \boldsymbol{\theta}, K_i = 1\right) \cong \hat{F}(\widehat{eCV}_i | \boldsymbol{y}_i, \boldsymbol{\theta}, K_i = 1) \tag{12}$$

With $\hat{F}$ being the ECDF of eCV_i. Equation (18) is analogous to a p-value. The larger the value of $\widehat{eCV}_i$, the larger the value of (18).

Second, we apply the Bayes theorem and the chain rule of probability to connect (18) with $\Pr(K_i|\boldsymbol{\theta},\mathbf{X})$:

$$\Pr(K_i = 1|\boldsymbol{\theta},\mathbf{X}) \propto$$
$$F\left(\widehat{eCV}_i\middle|\boldsymbol{y}_i,\boldsymbol{\theta},K_i = 1\right)\Pr(\boldsymbol{y}_i|\boldsymbol{\theta},K_i = 1)\Pr(\boldsymbol{\theta}|K_i = 1)\Pr(K_i = 1) \quad (19)$$

In a fully Bayesian treatment of (19), prior distributions for $\boldsymbol{\theta}$,K_i and their hyperparameters (e.g., π) might be used. For eCV, however, we assume a somewhat radical "Empirical Bayes" (EB) (Morris, 1983) method, where we set $\left(\widehat{\boldsymbol{\theta}}\middle|K_i = 1\right) \approx \left(\hat{\pi} = 0.5,\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\Sigma}}\right)$. The quantities $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are, respectively, the vector of means and the variance-covariance matrix across features. The values of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are assumed as good proxies for the parameters of prior distribution of reproducible features.

Third, EB estimates are plugged into equation (1) to randomly sample 10,000 values of $\boldsymbol{y}_i$ ($\tilde{\boldsymbol{y}}_i$). eCV is calculated in every $\tilde{\boldsymbol{y}}_i$ to obtain $\widetilde{eCV}_i$ values. The ECDF is used to obtain $\hat{F}$. Lastly, $\widehat{eCV}_i$ is plugged in $\hat{F}$ to obtain an estimate of (19).

2.3 Performance assessment

2.3.1 Defining a common reproducibility index across methods

Methods were compared using estimates of $\Pr(K_i = 0|\boldsymbol{\theta},\mathbf{X})$ as reproducibility index ($\widehat{\Pr}(K_i = 0|\boldsymbol{\theta},\mathbf{X})$). This index can be interpreted as the probability of an omic feature belonging to the irreproducible class. In the case of IDR and its extensions, gIDR and mIDR, the value $\Pr(K_i = 0|\boldsymbol{\theta},\mathbf{X})$ was approximated by the local IDR score (Li et. al 2011). For eCV, $\Pr(K_i|\boldsymbol{\theta},\mathbf{X})$ was approximated using parametric bootstrap.

276 *2.3.1 Performance quantification*

277 In all scenarios, the range of $\widehat{\Pr}\,(K_i = 0|\boldsymbol{\theta}, \mathbf{X})$ was split into 20 thresholds accommodated
278 in the vector $\boldsymbol{\tau} = [\tau_1, \tau_2, \ldots, \tau_{20}]$. The thresholds covered the range between 0 and 100%,
279 with each value τ_m splitting features into reproducible or not. Each threshold was
280 calculated as $\tau_m = \frac{100}{(1+e^{(10-m)})}$, where m is an integer from 1 to 20. The individual values
281 of the true state of each feature i (K_i) were imposed in each type of scenario. For
282 simulations, the value of K_i was a priori determined in each simulated data set, depending
283 on the value of π. For the experimental data sets, the value of K_i was imposed based on
284 biological knowledge from each system (Table 2). In all scenarios, the "true" feature state
285 was used to estimate true positives rate (TPR), showing true reproducible signal, and true
286 negative rates (TNR), showing true irreproducible noise.

287 *2.3.2 Inference and summary of performance values*

288 TPR and TNR across $\boldsymbol{\tau}$ values were calculated with the R package *pROC* (Robin *et al.*,
289 2011). For simulation scenarios, performance values were summarized in terms of
290 median and 90% confidence intervals across simulations. Performance values in
291 experimental data were summarized by the median and 90% confidence intervals across
292 2,000 bootstrap samples.

293 **2.4 Software implementation and data availability**

294 All methods and simulation scenarios are implemented as functions of the R package
295 *eCV.* All the software required to generate the results presented here are preinstalled in
296 a publicly available Docker image *ecv_results*. The image can be retrieved via *docker pull*
297 *ecvpaper2024/ecv_results:latest*. The miR-eCLIP data from transfections are available

298 from the GEO accession GSE198250 (miR-1 transfected samples, GSM5942220-1, miR-
299 124 transfected samples GSM5942222-3). The RBFOX2 eCLIP data are available from
300 the GEO accession GSE248884 (samples GSM7921694-701). The miR-eCLIP data from
301 wild-type K562 cells are available upon request.

302 ***3 RESULTS***

303 The four reproducibility assessment methods (**IDR**, **gIDR**, **mIDR**, and **eCV**) were used to
304 discriminate between reproducible and irreproducible features. Performance of each
305 method was evaluated in simulated and experimental data sets. In the case standard IDR
306 was used in settings with more than two replicates, only the first two were used. An
307 increasing number of features can significantly increase the computational costs of methods
308 in the IDR family. Although eCV is less affected by the total number of features, we narrowed
309 the analysis to only chromosome one to maintain consistency across comparisons.
310 Therefore, when comparing scenarios, we prioritized performance metrics based on
311 proportions to minimize the effect of a raw number of features. The instructions for
312 generating results using simulated data can be located in Supplementary File S1.
313 Additionally, instructions for generating results from the experimental data for miR-eCLIP
314 transfection, miR-eCLIP wild-type data, and RBFOX2-eCLIP are provided in
315 Supplementary Files S2, S3, and S4, respectively.

316 **3.1 Methods performance in simulated scenarios**

317 *3.1.1: Simulated Scenario 1: Many highly correlated, truly reproducible features exist.*

318 This scenario represented simulations where many highly correlated features were
319 reproducible (Figure 1). Performance of the IDR family in this scenario was more robust
320 to varying reproducibility thresholds τ than eCV, in both true positive (TPR) and true

negative rates (TNR). The method gIDR was consistently high in both TPR and TNR, particularly for increasing number of replicates r. The performance of mIDR, on the other hand, was more variable, with TNR increasing with r while TPR reaching lower values than the ones achieved with standard IDR. Lastly, the performance of method eCV marginally increased with r but did not reach the TPR performance achieved by IDR and its extensions. Both eCV and mIDR, however, achieved competitive performance at $\boldsymbol{\tau}$'s values between 50-75%

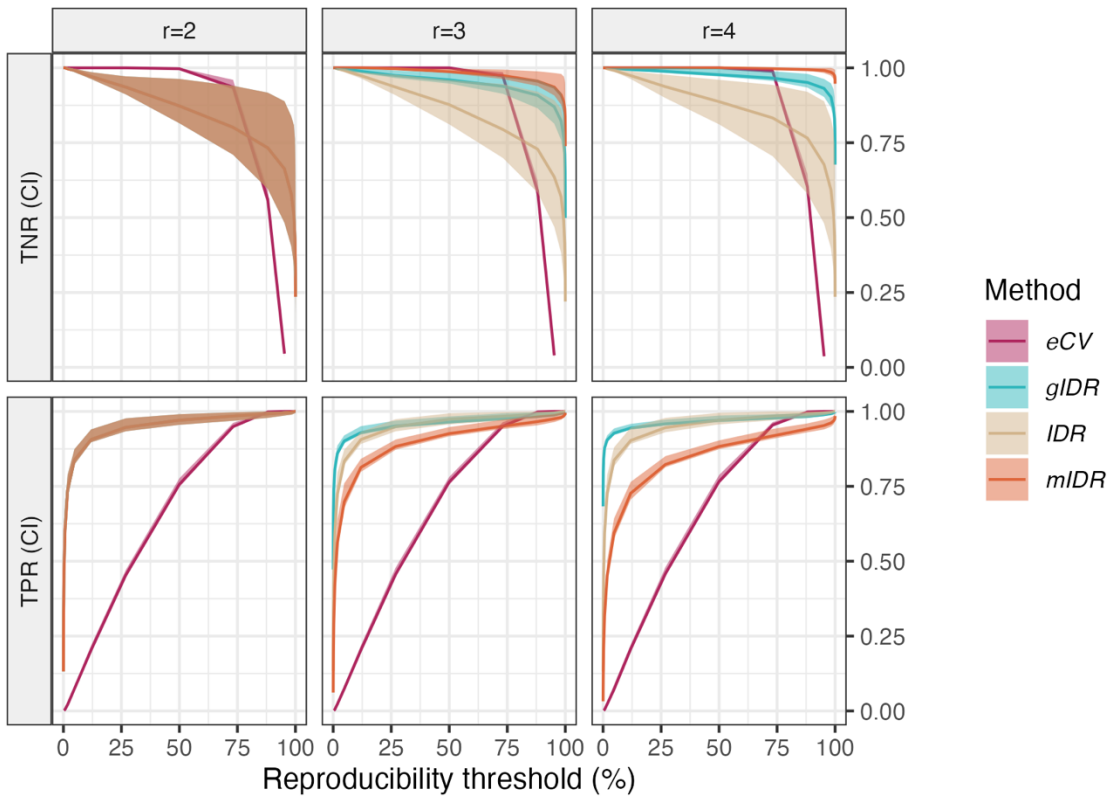


Figure 1: Methods performance in simulated scenario 1. *Many strongly correlated features of genuine signal*. Row panels are median and 90% credible intervals (**CI**) of true negative rate (**TNR**, top) and true positive rate (**TPR**, bottom) rates across 100 simulated

data sets. Column panels are varying number of replicates (r = 2, 3, or 4). The Y-axis is TPR or TNR values and the X-axis is the threshold on the reproducibility indices.

3.1.2: Simulated Scenario 2: Few weakly correlated truly reproducible features

This scenario represented situations where most features belonged to the irreproducible class and were lowly correlated (Figure 2). The overall performance values in this scenario were like the ones in scenario 1, with some differences. For example, gIDR gained even more accuracy over IDR (through increases in TNR values), but not power (no significant difference between IDR and gIDR's TPR values). The performance of eCV was once more overly sensitive to variations in thresholds $\boldsymbol{\tau}$. However, eCV's performance steadily increased with r, reaching competitive values at $\boldsymbol{\tau} \approx 50\%$. The method mIDR had marginally higher TNR values compared to IDR, but its TPR values significantly decreased with the number of replicates r.

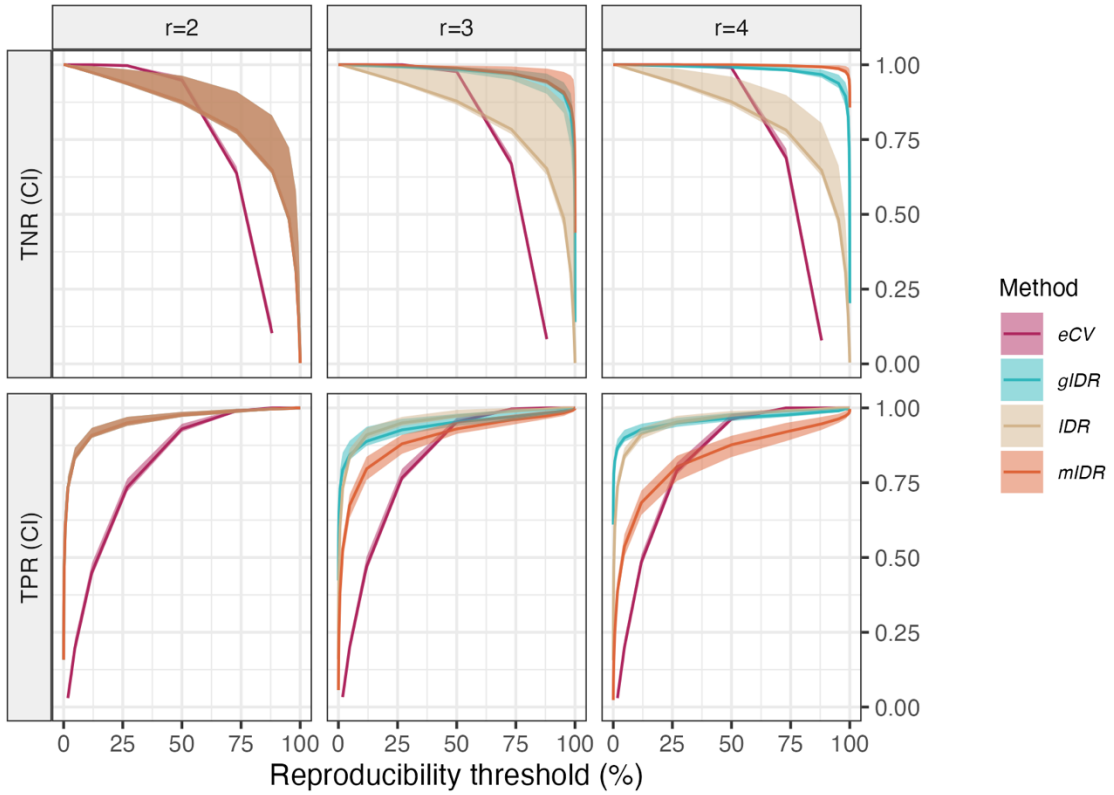


Figure 2: Methods performance in simulated scenario 2. *Few lowly correlated features of genuine signal*. Row panels are median and 90% credible intervals (**CI**) of true negative rate (**TNR**, top) and true positive rate (**TPR**, bottom) rates across 100 simulated data sets. Column panels stand for varying number of replicates (*r* = 2, 3, or 4). The Y-axis is TPR or TNR values and the X-axis is the threshold on the reproducibility indices.

3.1.3: Simulated Scenario 3: Few and highly correlated truly reproducible features.

The final simulated scenario represented situations with an even smaller number of strongly correlated reproducible features (Figure 3). In this scenario, eCV had the best overall performance across number of replicates r, achieving high TPR and TNR even for smaller thresholds $\boldsymbol{\tau}$.. Contrary to the earlier simulated scenarios, the performance of gIDR became severely affected by the increasing number of replicates r. Method mIDR,

on the other hand, systematically improved with r, achieving the best TNR values at $r = 4$.

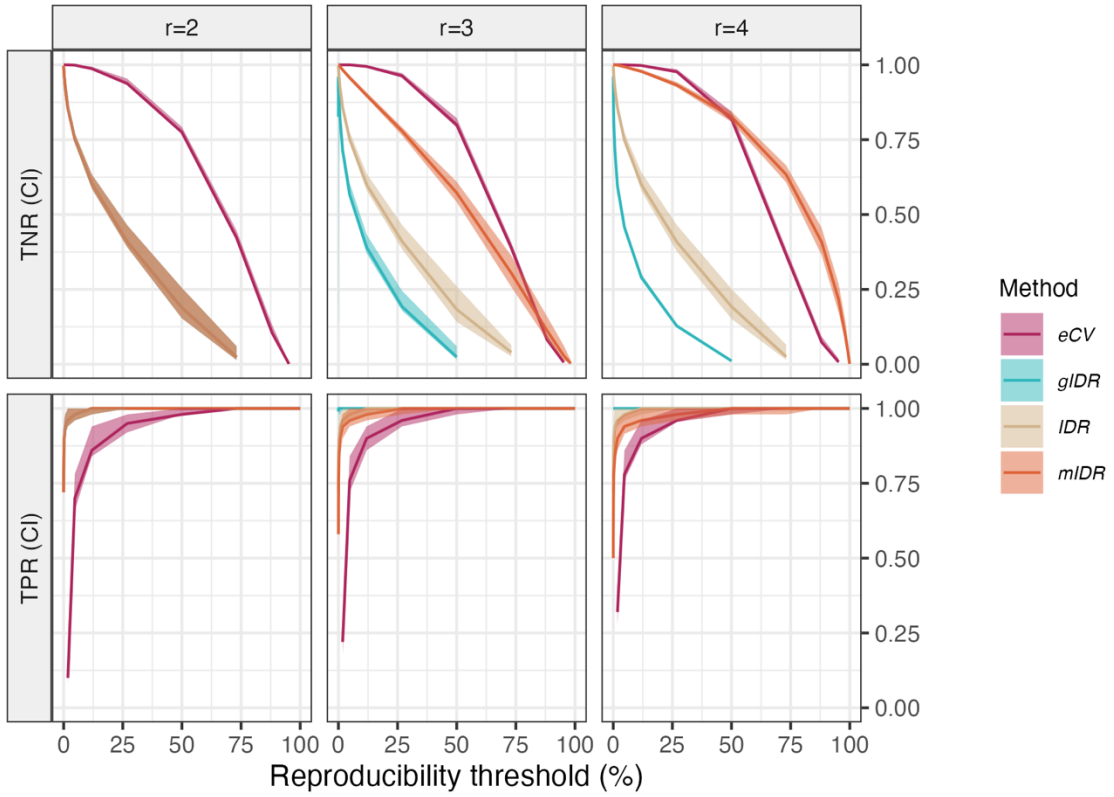


Figure 3: Methods performance in simulated scenario 3. *Few strongly correlated features of genuine signal*. Row panels stand for median and 90% credible intervals (**CI**) of true negative rate (**TNR**, top) and true positive rate (**TPR**, bottom) rates across 100 simulated data sets. Column panels are varying number of replicates (r = 2, 3, or 4). The Y-axis is TPR or TNR values and the X-axis is the threshold on the reproducibility indices.

3.2 Methods performance in experimental data

3.2.1: Experimental Scenario 1: miR-eCLIP data on transfected cells

The first experimental data set came from chimeric peaks data (peaks formed by chimeric reads coming from mRNA:miRNA interactions) obtained with miR-eCLIP on HEK293T cells. Peaks annotated with any of the two transfected miRNAs: *hsa-miR-124-3p* (miR-124) and *hsa-miR-1-3p* (miR-1), were used as proxies of truly reproducible features. Endogenous mRNA:miRNA chimeras, categorized as irreproducible, were considered as "true negatives." This selection was justified by the observation that the majority of endogenous chimeras in this dataset exhibited significantly lower intensities compared to those involving transfected miRNAs. Only n = 3,145). Of these, 1,560 were annotated with miR-124, while only 257 were annotated with miR-1. Confidence intervals at the 95% level suggested a high correlation of the peak intensities for both transfected miRNAs: [0.79, 0.83] for miR-124 and [0.55, 0.70] for miR-1. Since only two technical replicates were available in each transfection, only IDR and eCV were compared.

The results of the method comparisons in the miR-124 transfection resembled those obtained in simulated scenario 1 (i.e., many strongly correlated truly signal features). The results of the methods comparisons in the miR-1 transfection, on the other hand, resembled the ones in scenario 3 (i.e., small number of strongly correlated truly signal features). Although the global behavior of both IDR and eCV was like that in the simulations, eCV had significantly higher global performance than IDR, particularly for $\boldsymbol{\tau} \approx 50\%$.

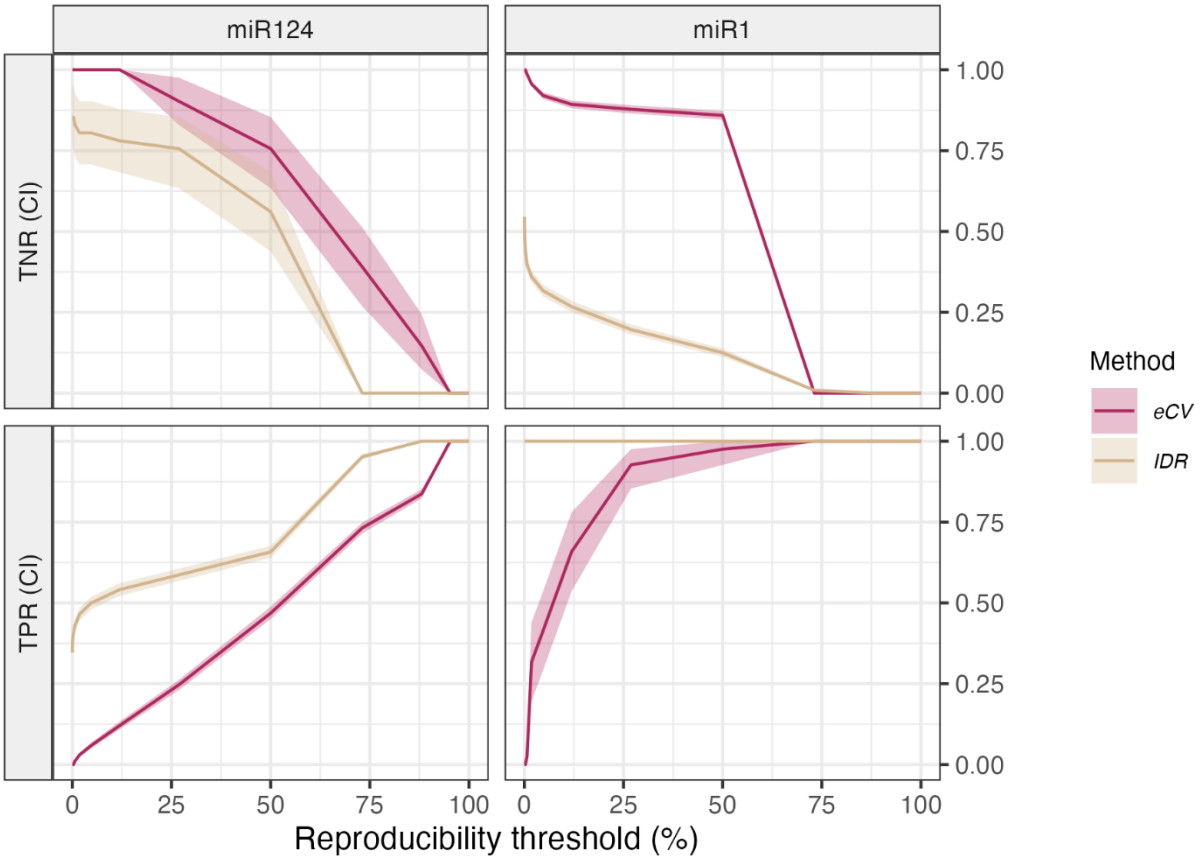


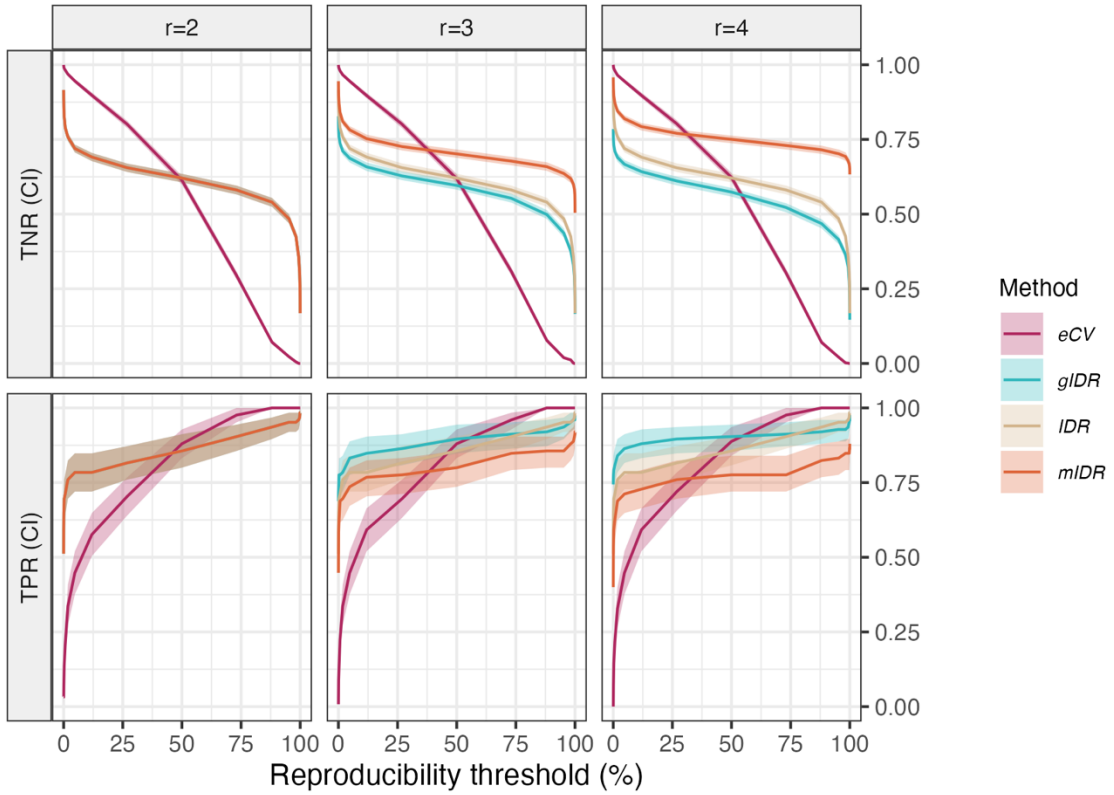
Figure 4: Methods performance in experimental scenario 1: *miR-eCLIP data from transfected cells.* Row panels are median and 90% credible intervals (**CI**) of true negative rate (**TNR**, top) and true positive rate (**TPR**, bottom) rates. Column panels are the transfections: *hsa-miR-124-3p* (**miR124**, left) and *hsa-miR-1-3p* (**miR1**, right). The Y-axis is TPR or TNR values and the X-axis is the threshold on the reproducibility indices.

3.2.2: Experimental Scenario 2: miR-eCLIP data from wild-type cells.

The second experimental data set also came from chimeric peaks data obtained with miR-eCLIP, but this time from two, three, and four replicates from wild-type K562 cells. Only chromosome 1 peaks were used (n = 1,858). Contrary to the earlier scenario, this one consisted of only endogenous chimeras. Therefore, only significantly enriched peaks

396 with positive seed matches on the 3' UTR were used as proxies of "truly reproducible"
397 features. Although the peaks being "truth" were strongly correlated within groups of genes
398 targeted by 13 miRNA species involved in those peaks (~0.9), the correlations between
399 groups miRNA significantly dropped.

400 Results resembled the ones in simulated scenario 2 (i.e., few lowly correlated truly
401 reproducible features). Like in scenario 2, gIDR's TNR values improved with increasing
402 number of replicates r, while TPR values were not significantly different than IDR's. Also,
403 in agreement with simulated scenario 2, eCV varied more quickly with thresholds $\boldsymbol{\tau}$ than
404 the other methods, while also steadily improving with the number of replicates. As before,
405 eCV matched the performance of every other method when $\boldsymbol{\tau} \approx 50\%$, except for mIDR's
406 TNR. The performance of mIDR also mirrored simulated scenario 2, except that here, its
407 TNR values were significantly higher than any other method.



408

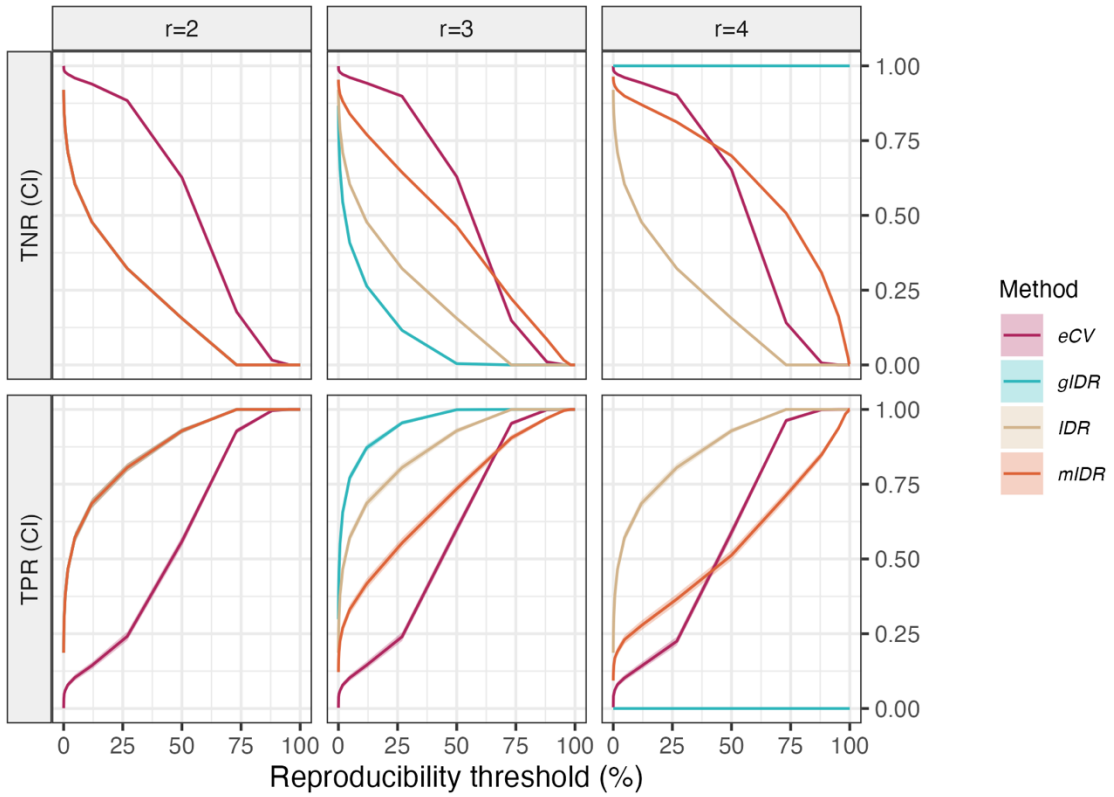
409 **Figure 5: Methods performance in experimental scenario 2:** *miR-eCLIP data from*
410 *wild-type cells.* Row panels are median and 90% credible intervals (**CI**) of true negative
411 rate (**TNR**, top) and true positive rate (**TPR**, bottom) rates. Column panels are varying
412 number of replicates (*r* = 2, 3, or 4). The Y-axis is TPR or TNR values and the X-axis is
413 the threshold on the reproducibility indices.

414 *3.2.3: Experimental Scenario 3: RBFOX2 eCLIP data on wildtype cells*

415 The third and last experimental data set came from peak data derived from an RBFOX2
416 eCLIP on HEK293T cells for 2, 3, and 4 replicates. Peaks enriched with the RBFOX2
417 binding motif and mapping onto the 3' UTR were proxies for "truly reproducible" features.
418 Only chromosome 1 peaks were used (n = 24,703). Of these, only 5,173 fulfilled both
419 conditions, with approximately 20% of the peaks having the RBFOX2 motif and

420 approximately 13% mapping onto the 3' UTR. Correlations among these peaks were also
421 highly significant: confidence intervals at 95% = [0.60, 0.63].

422 Results in this scenario resembled the ones for simulated scenario 3, distinguished by
423 only a handful of highly correlated truly reproducible features. As for simulated scenario
424 3, eCV reached higher TPR values with increasing r while systematically keeping higher
425 TNR, particularly at $\boldsymbol{\tau}$ around 50%. Similarly, mIDR improved its TNR values for increasing
426 r. However, contrary to simulation results, its TPR seemed to decrease. Also, consistently
427 with simulated scenario 3, gIDR performance was severely affected, particularly for
428 increasing r.



429

430 **Figure 6: Methods performance in experimental scenario 3**. *RBFOX2 eCLIP data*
431 *from wildtype cells.* Row panels are median and 90% credible intervals (**CI**) of true

negative rate (**TNR**, top) and true positive rate (**TPR**, bottom) rates. Column panels are varying number of replicates (*r* = 2, 3, or 4). The Y-axis is TPR or TNR values and the X-axis is the threshold on the reproducibility indices.

4 DISCUSSION

The escalating capacity for multiplexing and diminishing sequencing costs make it possible to work with increasing sample sizes, thereby augmenting the quality of high-throughput data by raising power and precision (Chen et al., 2012; Li et al., 2017). Nevertheless, reproducibility analysis is still needed to ensure valid results. This becomes a non-trivial problem when using standard techniques, such as irreproducible discovery rate (IDR), on more than two replicates (Yang et al., 2014). We propose three novel methods to overcome this limitation as alternatives to the traditional IDR. Results from both simulations and experimental data suggest that the methods presented here can outperform IDR in several relevant biological scenarios.

Both simulated and experimental data scenarios were chosen to show some semblance to each other. For example, the chimeric peaks in experimental scenario 1 (cells transfected with miR-124 or miR-1) annotated with the respective transfected miRNA resulted in significantly high correlations. These high correlations and the fact that miR-124 had more than 83% of annotated peaks, while miR-1 had about 14%, could explain the strong resemblance between this data set and simulation scenarios 1 and 3, standing for many and few strongly correlated truly reproducible features, respectively. A similar explanation can be applied to the behaviors between simulated scenarios two and three and the second miR-eCLIP and RBFOX2 eCLIP data sets.

454 Of all the methods compared, the general IDR (gIDR) was particularly powerful at
455 detecting truly reproducible features while discarding truly irreproducible ones in many of
456 the scenarios tested. This behavior manifested in two opposite situations: many highly
457 correlated reproducible features (e.g., simulation scenario 1) and few lowly-correlated
458 reproducible features (e.g., simulation scenario 2). Unfortunately, the performance of
459 gIDR significantly dropped when even fewer but highly correlated reproducible features
460 were present (e.g., simulation scenario 3). In those cases, gIDR underperformed with an
461 increasing number of replicates. One compelling explanation for these behaviors is the
462 high sensitivity of multivariate normal components to the misspecification of initial
463 parameter values in combination with ill-conditioned variance-covariance matrices (Gray,
464 1994; Lo, 2011). Consider for example the determinant of the variance-covariance matrix
465 of the reproducibility component $\mathbf{\Sigma}_1$, $\left|\sigma_1^2(1-\rho_1)\left[\mathbf{I}+\frac{\rho_1}{1-\rho_1}\mathbf{J}\right]\right| = (\sigma_1^2)^r(1+(r-1)\rho_1)(1-$
466 $\rho_1)^{r-1}$, where as before, σ_1^2 and ρ_1 the reproducible class variance and correlation. If,
467 say $\sigma_1^2 = 1$ and $\rho_1 = 0.9$, then the determinant $|\mathbf{\Sigma}_1| = \frac{9(1+r)}{10^r}$ quickly approaches zero as
468 r grows, making $\mathbf{\Sigma}_1$ near singular. This, combined with near-boundary values of the mixing
469 parameter π (e.g., 0.05, as in simulated scenario 3), might artificially push features into
470 one mixture components, regardless of the actual reproducibility state. Potential solutions
471 to these issues are application of regularization during estimations of $\mathbf{\Sigma}_1$ (Liu and Palomar,
472 2019) and preoptimized initial values of parameters parameter values (Panić *et al.*, 2020).
473 Meta IDR (mIDR) performance varied more than gIDR, being particularly high in
474 simulated scenario 3. Although the experimental RBFOX2 eCLIP scenario resembled
475 simulated scenario 3, the method did not perform as well, particularly in terms of TPR.
476 Nevertheless, it seems that mIDR would increase its performance with an increasing

477 number of replicates and be particularly good in situations where only a few features are
478 truly reproducible. Although the performance of mIDR would increase with r, since
479 traditional IDR is applied in every pairwise combination, the number of total comparisons
480 would increase quadratically. Although the calculation of IDR in every pair of replicates
481 can be achieved in parallel, it is also worth mentioning potentially high computational
482 costs if several replicates are jointly analyzed.
483 The global performance of eCV was variable across scenarios. Although with the highest
484 sensitivity to the choice of the threshold on the reproducibility values, eCV was
485 undoubtedly the more consistent method across varying numbers of replicates. eCV was
486 also less potent than IDR when most features were reproducible and highly correlated.
487 These facts point to eCV being a more stringent method than IDR, potentially due to the
488 heavy prior information integrated into the model via Empirical Bayes. Nevertheless, eCV
489 still detected the most reproducible features (e.g., those with higher feature intensities,
490 Supplementary Figure S1) while calling fewer false positives. In addition, eCV also had
491 significantly more global accuracy, most clear in the miR-124 miR-eCLIP experiment. As
492 opposed to gIDR, eCV would also have higher performance in scenarios with fewer
493 reproducible features.
494 Further research on the applicability of a fully Bayesian treatment of eCV could alleviate
495 the potential stringency of the method. Nevertheless, due to the low sample sizes typically
496 used in high-throughput experiments, a strong prior via EB, like the one used in the *limma*
497 suit of models (Ritchie et al., 2015) could augment power. The ability to detect highly
498 reproducible features, combined with its high TNR across different scenarios, makes eCV
499 a competitive method, particularly when a threshold of 50% was used. This threshold is

also compelling since it matches the intuition of calling features as reproducible; their values surpass what is expected from entirely random behavior.

All the methods were compared on the same scale. In the case of IDR and its extensions, this reduced to working with the so-called local IDR. In the case of eCV, this involved using the probability of the estimated eCV being significantly larger than values expected for reproducible features. Although a method for FDR control via eCV was not tested, a similar approach to the one used by IDR could be implemented. This would involve plugging the eCV's reproducibility indices into a multiple comparison approach, like the one adopted by IDR following Sun and Cai (2007).

We have shown that multivariate alternatives to standard IDR might perform highly, but choosing the best model still is a non-trivial issue. Although the results shown here supply some guidelines, due to most biological problems not having a clear sign of "truth," deciding when to use one or another method is not trivial. Therefore, deciding which method would perform the best would depend on various aspects, such as what type of statistical error one wishes to control the most. Due to its robustness to varying sample sizes, stable behavior, and high accuracy, we favor eCV as a suitable high-performance method across many contrasting scenarios.

Conflict of interest

Eclipsebio funded and conducted the research presented in this paper. The authors are paid employees of Eclipsebio but have no financial interests or personal relationships that could potentially bias the results or interpretation of the data.

521 ***REFERENCES***

522 Arya,A.D. *et al.* (2014) RBFOX2 protein domains and cellular activities. *Biochem. Soc.*
523 *Trans.*, **42**, 1180–1183.

524 Chen,Y. *et al.* (2011) MM-ChIP enables integrative analysis of cross-platform and
525 between-laboratory ChIP-chip or ChIP-seq data. *Genome Biol.*, **12**, R11.

526 Chen,Y. *et al.* (2012) Systematic evaluation of factors influencing ChIP-seq fidelity. *Nat.*
527 *Methods*, **9**, 609–614.

528 Dwyer,P.S. (1967) Some Applications of Matrix Derivatives in Multivariate Analysis. *J. Am.*
529 *Stat. Assoc.*, **62**, 607–625.

530 Frankish,A. *et al.* (2022) GENCODE: reference annotation for the human and mouse
531 genomes in 2023. *Nucleic Acids Res.*, **51**, D942–D949.

532 Genest,C. and Nešlehová,J. (2013) Copulas and Copula Models. In, *Encyclopedia of*
533 *Environmetrics*. John Wiley & Sons, Ltd.

534 Gray,G. (1994) Bias in misspecified mixtures. *Biometrics*, **50**, 457–470.

535 Gu,S. *et al.* (2009) The biological basis for microRNA target restriction to the 3'
536 untranslated region in mammalian mRNAs. *Nat. Struct. Mol. Biol.*, **16**, 144–150.

537 Haecker,I. *et al.* (2012) Ago HITS-CLIP Expands Understanding of Kaposi's Sarcoma-
538 associated Herpesvirus miRNA Function in Primary Effusion Lymphomas. *PLOS*
539 *Pathog.*, **8**, e1002884.

540 Heinz,S. *et al.* (2010) Simple combinations of lineage-determining transcription factors
541 prime cis-regulatory elements required for macrophage and B cell identities. *Mol.*
542 *Cell*, **38**, 576–589.

543 Landt,S.G. *et al.* (2012) ChIP-seq guidelines and practices of the ENCODE and
544 modENCODE consortia. *Genome Res.*, **22**, 1813–1831.
545 Li,C.-I. *et al.* (2017) Power and sample size calculations for high-throughput sequencing-
546 based experiments. *Brief. Bioinform.*, **19**, 1247–1255.
547 Li,Q. *et al.* (2011) Measuring reproducibility of high-throughput experiments. *Ann. Appl.*
548 *Stat.*, **5**.
549 Liu,J. and Palomar,D.P. (2019) Regularized robust estimation of mean and covariance
550 matrix for incomplete data. *Signal Process.*, **165**, 278–291.
551 Lo,Y. (2011) Bias from misspecification of the component variances in a normal mixture.
552 *Comput. Stat. Data Anal.*, **55**, 2739–2747.
553 Manakov,S.A. *et al.* (2022) Scalable and deep profiling of mRNA targets for individual
554 microRNAs with chimeric eCLIP. 2022.02.13.480296.
555 Morris,C.N. (1983) Parametric Empirical Bayes Inference: Theory and Applications. *J.*
556 *Am. Stat. Assoc.*
557 Mundade,R. *et al.* (2014) Role of ChIP-seq in the discovery of transcription factor binding
558 sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell*
559 *Cycle*, **13**, 2847–2852.
560 Nicholson,G. and Holmes,C. (2017) A note on statistical repeatability and study design
561 for high-throughput assays. *Stat. Med.*, **36**, 790–798.
562 Panić,B. *et al.* (2020) Improved Initialization of the EM Algorithm for Mixture Model
563 Parameter Estimation. *Mathematics*, **8**, 373.

564 Revilla-i-Domingo,R. *et al.* (2012) The B-cell identity factor Pax5 regulates distinct
565 transcriptional programmes in early and late B lymphopoiesis. *EMBO J.*, **31**, 3130–
566 3146.
567 Robin,X. *et al.* (2011) pROC: an open-source package for R and S+ to analyze and
568 compare ROC curves. *BMC Bioinformatics*, **12**, 77.
569 Rozowsky,J. *et al.* (2009) PeakSeq enables systematic scoring of ChIP-seq experiments
570 relative to controls. *Nat. Biotechnol.*, **27**, 66–75.
571 Savits,T.H. (1994) On integration, substitution and the probability integral transform. *Stat.*
572 *Probab. Lett.*, **21**, 173–179.
573 Shelby,L.B. and Vaske,J.J. (2008) Understanding Meta-Analysis: A Review of the
574 Methodological Literature. *Leis. Sci.*, **30**, 96–110.
575 Soutschek,M. *et al.* (2022) scanMiR: a biochemically based toolkit for versatile and
576 efficient microRNA target prediction. *Bioinformatics*, **38**, 2466–2473.
577 Sun,W. and Cai,T.T. (2007) Oracle and Adaptive Compound Decision Rules for False
578 Discovery Rate Control. *J. Am. Stat. Assoc.*, **102**, 901–912.
579 Tsukahara,H. (2005) Semiparametric estimation in copula models. *Can. J. Stat.*, **33**, 357–
580 375.
581 Van Nostrand,E.L. *et al.* (2016) Robust transcriptome-wide discovery of RNA-binding
582 protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods*, **13**, 508–514.
583 Xu,J. *et al.* (2016) Global Analysis of Expectation Maximization for Mixtures of Two
584 Gaussians. In, *Advances in Neural Information Processing Systems*. Curran
585 Associates, Inc.
586

587